Universal Quantization of Parametric Sources has Redundancy $\frac{k}{2}\frac{\log n}{n}$

P.A. Chou*, M. Effros†, and R.M. Gray‡

*Xerox Palo Alto Research Center, 3333 Coyote Road, Palo Alto, CA 94304
†Dept. of Electrical Engineering, 116-81, Caltech, Pasadena, CA 91125
‡Information Systems Laboratory, Stanford University, Stanford, CA 94305-4055

Abstract — **Let $\{X_i\} \sim P_\theta$, $\theta \in \Lambda \subseteq \mathbb{R}^k$. Rissanen has shown that there exist universal noiseless codes for $\{X_i\}$ with per-letter rate redundancy as low as $\frac{k}{2}\frac{\log n}{n}$, where n is the blocklength and k is the number of source parameters. We derive an analogous result for universal quantization: for any given Lagrange multiplier $\lambda > 0$, there exist universal fixed-rate and variable-rate quantizers with per-letter Lagrangian redundancy (i.e., distortion redundancy plus λ times the rate redundancy) as low as $\lambda\frac{k}{2}\frac{\log n}{n}$.**

Let $\{X_i\}$ be a stationary ergodic random process over alphabet $\mathcal{X}$ with process measure P_θ, $\theta \in \Lambda \subseteq \mathbb{R}^k$, and let $C^n = \beta^n \circ \alpha^n$ be a length-n quantizer with encoder $\alpha^n : \mathcal{X}^n \to \mathcal{S}$ and decoder $\beta^n : \mathcal{S} \to \mathcal{Y}^n$, where $\mathcal{S} = \{s_1, \ldots, s_M\} \subseteq \{0,1\}^*$ is some binary prefix code and $\mathcal{Y}$ is the reproduction alphabet. Let $d(x^n, y^n) = \sum_i d(x_i, y_i)$ be a single-letter fidelity criterion and let $|s|$ denote the length of the binary string s. The nth order operational distortion-rate function for $\{X_i\}$ is defined

$$\hat{D}_\theta^n(R) = \inf_{C^n} \left\{ \frac{1}{n} E_\theta d(X^n, C^n(X^n)) : \frac{1}{n} E_\theta |\alpha^n(X^n)| \le R \right\},$$

where the infimum is over either fixed-rate or variable-rate quantizers with blocklength n, as appropriate. The support functional of $\hat{D}_\theta^n(R)$ is defined

$$\hat{L}_\theta^n(\lambda) = \inf_{C^n} \left[\frac{1}{n} E_\theta d(X^n, C^n(X^n)) + \lambda \frac{1}{n} E_\theta |\alpha^n(X^n)| \right],$$

where $\lambda > 0$ is a Lagrange multiplier.

We show that there exists a universal sequence of fixed-rate or variable-rate quantizers $\{C^n\}$ such that the per-letter Lagrangian

$$\ell_\theta(\lambda, C^n) = \frac{1}{n} E_\theta d(X^n, C^n(X^n)) + \lambda \frac{1}{n} E_\theta |\alpha^n(X^n)|$$

converges to the support functional $\hat{L}_\theta^n(\lambda)$ as $\lambda\frac{k}{2}\frac{\log n}{n}$ for every $\theta \in \Lambda \subseteq \mathbb{R}^k$. To be precise, assume that for every θ, λ, and n, $\hat{L}_\theta^n(\lambda)$ is achieved by some C^n, say $C_{\theta,\lambda}^n$. Then define

$$\Delta_\lambda^n(\theta||\hat{\theta}) = \ell_\theta(\lambda, C_{\hat{\theta},\lambda}^n) - \hat{L}_\theta^n(\lambda)$$

to be the divergence between the Lagrangian performance of the quantizer matched to $\hat{\theta}$ and the quantizer matched to θ, with respect to θ. We have the following:

Theorem 1 *Let Λ be a subset of $\mathbb{R}^k$ (bounded if we are considering fixed-rate coding but possibly unbounded otherwise). Suppose that for each θ, λ, and n there exists a code $C_{\theta,\lambda}^n$ achieving the support functional $\hat{L}_\theta^n(\lambda)$. Suppose also that the corresponding divergence $\Delta_\lambda^n(\theta||\hat{\theta})$ is locally quadratic such that for each θ and λ there exists a neighborhood $S_{\theta,\lambda}$ of θ and a constant $m_{\theta,\lambda}$ such that $\Delta_\lambda^n(\theta||\hat{\theta}) \le m_{\theta,\lambda}||\theta - \hat{\theta}||^2$ for all $\hat{\theta} \in S_{\theta,\lambda}$ and for all n. Then for each λ there exists a weakly minimax universal sequence of codes $\{C^n\}$ such that for all θ*

$$\ell_\theta(\lambda, C^n) - \hat{L}_\theta^n(\lambda) \le \lambda \frac{k}{2}\frac{\log n + c_{\theta,\lambda}}{n}.$$

If Λ is bounded, and $S_{\theta,\lambda}$ and $m_{\theta,\lambda}$ do not depend on θ, then neither does $c_{\theta,\lambda}$, and the sequence $\{C^n\}$ is strongly minimax universal.

Proof: Fix λ. Construct $C^n = \beta^n \circ \alpha^n$ as follows. For each $n \ge 1$, partition $\mathbb{R}^k$ into a grid of hypercubes $\{A_i^n : i = 1, 2, \ldots\}$ each with side $1/\lceil n^{1/2} \rceil$, such that $\{A_i^n : i = 1, 2, \ldots\}$ refines $\{A_j^1 : j = 1, 2, \ldots\}$. For each hypercube A_i^n that intersects Λ, choose a representative $\hat{\theta}_i^n \in A_i^n \cap \Lambda$ and its matching quantizer $C_i^n = C_{\hat{\theta}_i^n,\lambda}^n$. Then define the encoder α^n to map x^n to the string $s = s_i' s_i'' s_i'''$ where s_i' represents the unit hypercube A_j^1 containing A_i^n, (which can be a fixed-length string if Λ is bounded), s_i'' represents the hypercube A_i^n indexed within A_j^1 (which is a fixed-length string with length $\log \lceil n^{1/2} \rceil^k$), and s_i''' is the string $\alpha_i^n(x^n)$ representing x^n using the quantizer C_i^n. The decoder maps s to the reproduction $y^n = \beta_i^n(s_i''')$. The index i is chosen to minimize the instantaneous Lagrangian $d(x^n, C_i^n(x^n)) + \lambda|s_i' s_i'' s_i'''|$. Thus

$$\begin{aligned} d(x^n, C^n(x^n)) + \lambda|\alpha^n(x^n)| &= \min_i d(x^n, C_i^n(x^n)) + \lambda|s_i' s_i'' \alpha_i^n(x^n)| \\ &\le d(x^n, C_j^n(x^n)) + \lambda|s_j' s_j'' \alpha_j^n(x^n)| \end{aligned}$$

for any particular j. Let j be the index of the cell A_j^n containing θ. Then dividing by n, taking expectations, and subtracting $\hat{L}_\theta^n(\lambda)$, we have

$$\begin{aligned} \ell_\theta(\lambda, C^n) - \hat{L}_\theta^n(\lambda) &\le \ell_\theta(\lambda, C_j^n) - \hat{L}_\theta^n(\lambda) + \frac{\lambda}{n}|s_j' s_j''| \\ &\le \Delta_\lambda^n(\theta||\hat{\theta}_j^n) + \frac{\lambda}{n}\left(b_\theta + \frac{k}{2}\log n\right), \end{aligned}$$

for some constant b_θ. By assumption, $\Delta_\lambda^n(\theta||\hat{\theta}) \le m_{\theta,\lambda}||\theta - \hat{\theta}||^2$ for all $\hat{\theta}$ in a neighborhood $S_{\theta,\lambda}$ of θ. Since $\hat{\theta}_j^n \to \theta$ with $||\theta - \hat{\theta}_j^n||^2 \le k/n$, there exists a constant $a_{\theta,\lambda}$ such that $\Delta_\lambda^n(\theta||\hat{\theta}_j^n) \le a_{\theta,\lambda} k/n$ for all n. Thus the theorem is proved with $c_{\theta,\lambda} = 2a_{\theta,\lambda}/\lambda + 2b_\theta/k$. □

A simple example of a source satisfying the conditions of the theorem is the following. Let $Z_1, Z_2, \ldots$ be an arbitrary real-valued stationary ergodic process with mean 0 and variance 1, and let $X_i = \sigma Z_i + \mu$. Then with $\theta = (\mu, \sigma) \in \Lambda \subseteq \mathbb{R}^2$, under the squared-error distortion measure and fixed-rate quantization of $\{X_i\}$, for all λ, n, θ, and $\hat{\theta}$, $\Delta_\lambda^n(\theta||\hat{\theta}) \le ||\theta - \hat{\theta}||^2$. Hence for any stationary source with unknown mean and variance in a bounded set, there exists a strongly minimax universal sequence of fixed-rate quantizers for which the nth order Lagrangian redundancy is at most $\lambda(k/2)(\log n + c)/n$, where $k = 2$.